Designing New Machine Learning Techniques

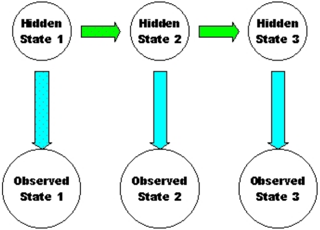
Presentation 2

Eli Ben-Michael Mentor: Dr. Kevin Chen

July 18, 2014

Outline

Hidden Markov Models



- Set of observations and hidden states
- Transition from hidden state to hidden state is a **Markov Process**
- Markov Condition: Given a set of random variables $\{H_t | t \in \mathbb{N}\}$, $P(H_{t+1} = h_{t+1} | H_t = h_t, \ldots, H_0 = h_0) = P(H_{t+1} = h_{t+1} | H_t = h_t)$

Hidden Markov Models

d = number of observed states
k = number of hidden states
h_t = hidden state at time t
x_t = observed state at time t

Parameters:

- $\vec{\pi} \in \mathbb{R}^k$: initial hidden state distribution
- $T \in \mathbb{R}^{d \times k}$: transition matrix
  - $P(h_{t+1} = i | h_t = j) = T_{i,j}$
- $O \in \mathbb{R}^{d \times k}$: emission matrix
  - $\mathbb{E}[x_t | h_t = j] = O\vec{e}_j$

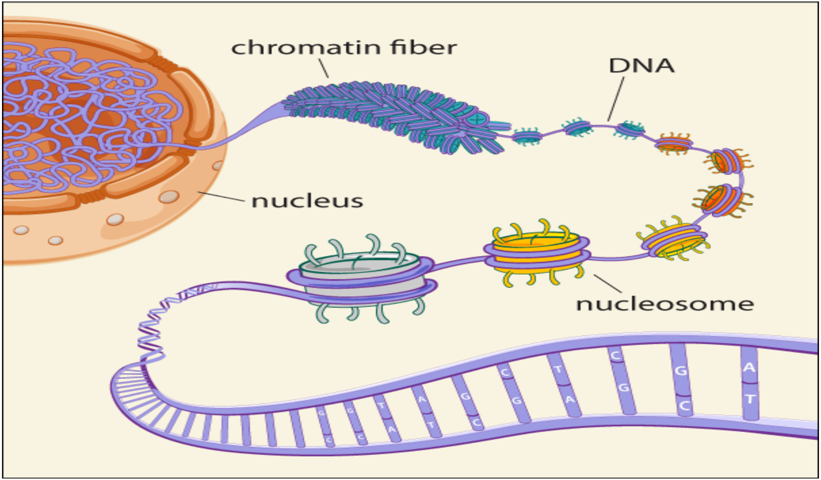
Chromatin

- Chromatin inside of nucleus hold genetic information
- Hundreds of slight modifications ("marks")

Question

Do chromatin marks correlate with functional elements in genome? If so, how?

Chromatin



Hidden Markov Models and Chromatin Marks

- Can use Hidden Markov Models to find the chromatin states based on which combinations of chromatin marks are most prevalent
- Observations: Chromatin Marks
- Hidden States: Chromatin States (associated with functional elements)

Machine Learning: Finding the parameters of the HMM

Outline

Tensors

Tensor

A **Real pth-order Tensor** $T \in \bigotimes^p \mathbb{R}^n$ is a p-way array where $T_{i_1,i_2,\ldots,i_p}$ is the element at position $i_1, i_2, \ldots, i_p$

Tensor product and Tensor power

For any vectors $\vec{v}, \vec{w} \in \mathbb{R}^n$, $\vec{v} \otimes \vec{w}$ denotes the **Tensor Product** of $\vec{v}$ and $\vec{w}$, and $\vec{v}^{\otimes p} = \vec{v} \otimes \vec{v} \otimes \vec{v} \otimes \ldots \otimes \vec{v} \in \bigotimes^p \mathbb{R}^n$ denotes the p-th **Tensor Power**

Example tensor product

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}, \vec{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}, \vec{v} \otimes \vec{w} = \begin{pmatrix} v_1 w_1 & v_1 w_2 \\ v_2 w_1 & v_2 w_2 \end{pmatrix}$$

Tensor Spectral Decomposition Problem

The problem

Given a 3rd order tensor $T \in \otimes^3 \mathbb{R}^n$ that is of the form

$$T = \sum_{j=1}^{n} \lambda_j \vec{v}_j^{\otimes 3}$$

where $\{\vec{v}\}$ is an orthonormal basis of $\mathbb{R}^n$ and $\lambda_i > 0 \forall i$. Can we (approximately) find all of the $\vec{v}_i, \lambda_i$ pairs?

Answer: Yes. Tensor Power Method from last presentation

Method of Moments

But how do we use Tensor Spectral Decomposition to recover the parameters of the HMM?

Method of Moments

Proposition

Through a linear transformation of the data we can extract transformed second and third moments, M_2, and M_3 such that:

$$M_2 = \sum_{i=1}^{k} \omega_i (O\vec{e_i})^{\otimes 2}$$

$$M_3 = \sum_{i=1}^{k} \omega_i (O\vec{e_i})^{\otimes 3}$$

where

$$\vec{\omega} = T\pi \in \mathbb{R}^k$$

T and M_3

Compare

$$T = \sum_{j=1}^{n} \lambda_j \vec{v}_j^{\otimes 3}$$

with

$$M_3 = \sum_{i=1}^{k} \omega_i (O\vec{e}_i)^{\otimes 3}$$

Two differences:

- T is rank n and $n \times n \times n$, M_3 is rank k and $d \times d \times d$ with $d > k$
- $\{\vec{v}_j\}$ are orthonormal, $\{O\vec{e}_i\}$ are not

Implementation

- Python
- NumPy and SciPy

Key Computation

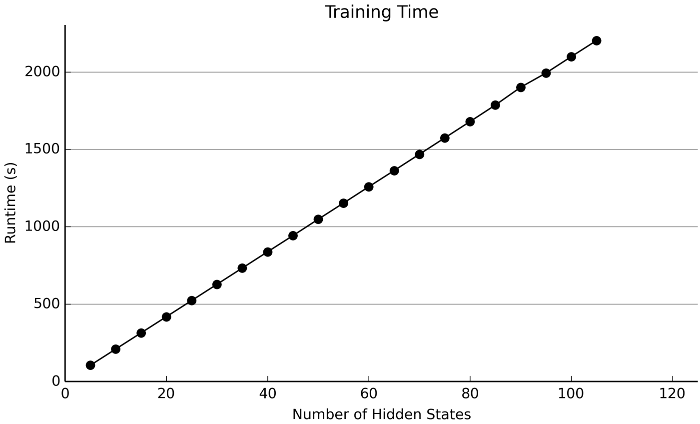
$$\phi_T(\vec{x}) = \sum_{i,j,k} T_{i,j,k} x_j x_k \vec{e}_i$$

- Have to do this many times
- Triple sum, takes time
- Rewrite as matrix multiplication:

$$\phi_T(\vec{x}) = \sum_i \vec{e}_i \vec{x}^\intercal T_i \vec{x}$$

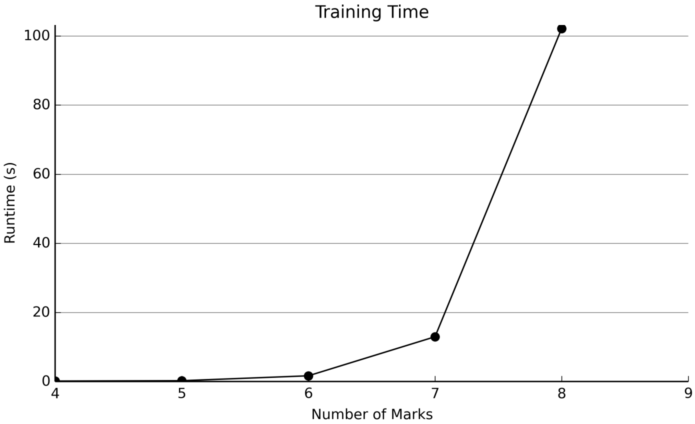
- Matrix multiplication is optimized in NumPy, speeds up algorithm 100 fold

Runtime



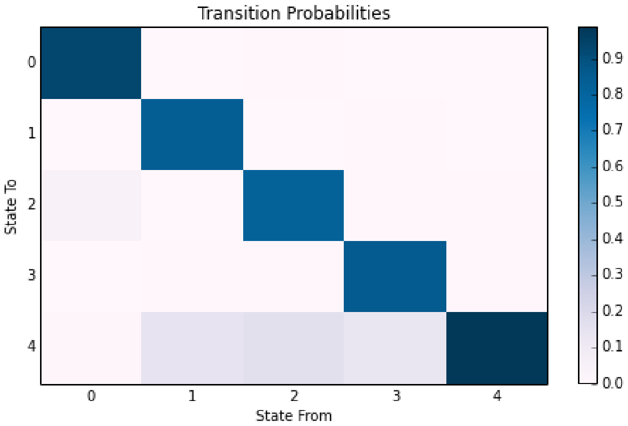
Training Time
Runtime (s)
0
500
1000
1500
2000
0
20
40
60
80
100
120
Number of Hidden States

Runtime

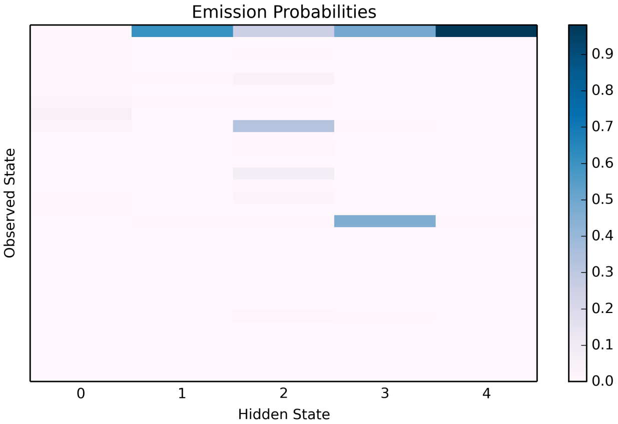


Training Time
100
80
60
40
20
0
Runtime (s)
4
5
6
7
8
9
Number of Marks

Sample Output



Sample Output



References

Anandkumar, A., Ge, R. Hsu, D., Kakade, S. and Telgarsky, M. *Tensor Decompositions for Learning Latent Variable Models* arXiv:1210.7559 [cs.LG] 2014

Anandkuma, A., Foster, D., Hsu, D., Kakade, S., Liu, Y. emphA Spectral Algorithm for Latent Dirichlet Allocation arXiv:1204.6703 [cs.LG] 2012

Baker, M. *Making sense of chromatin states. Nature Methods* 2011